Exercise 4

Machine Learning I

|  |  |
| --- | --- |
|  | 4A-1. |

|  |
| --- |
| Prerequisites |

|  |
| --- |
| Leibnitz rule for constants  Additionally, let at least one of the following conditions hold:   1. is measurable and nonnegative 2. is finite   then we can switch the order of integration according to Tonelli/Fubini, respectively.  Note: In our case, at least one of (i), (ii) is nearly always satisfied. Think about why.  Lastly, we need *Theorem 1* concerning uniform convergence from these notes:  http://www.math.ucla.edu/~tao/resource/general/131bh.1.03s/week45.pdf |

Let be the domain of .

Minimizing the loss cumulative loss for all **t** equals minimizing the loss for each separately. Note: Inside the interior integral, is constant. Let :

We can now solve for

According to the Leibnitz rule, this only holds for finite limits . To extend this proof to the infinite domain, we construct the sequence:

Because probabilities sum to one, if tends to infinity, vanishes for most The will not compensate that, as we required to be finite earlier.  
  
Accordingly, converges to uniformly. Due to , the functions converge uniformly to , with   
  
Spoken plainly, this means if we have an infinite domain , we can approximate the solution arbitrarily close by increasing .

Only because the are independent. In our situation this is the case, otherwise we would also have to integrate over all possibilities .

|  |  |
| --- | --- |
|  | 4A-2. |

|  |
| --- |
| Auxiliary calculation |

|  |
| --- |
| Using the Jacobian integral substitution, the area of an infinitesimal -dimensional volume element is  Above can be seen here:  <https://en.wikipedia.org/wiki/N-sphere#Spherical_coordinates>  Furthermore: We have for all because the squared length of a coordinate point is .  The surface area of an dimensional sphere with radius is denoted by  where denotes the volume of a sphere of radius . This can be seen here: <http://scipp.ucsc.edu/~haber/ph116A/volume_11.pdf>  There we can also find the identity (Eq.7 ): |

Ergo .

Now we are looking for the maximum density :

Because radii are non-negative, we have a maximum at .  
  
Now if we set we get

|  |  |
| --- | --- |
|  | 4A-3. |

Let

Derivative of with respect to :

Conversion to matrix/vector operations:

Generalized for all

Setting zero and solving for :

As usual, denotes the Penrose pseudo inverse.  
Because can be an arbitrary normative factor, it is also possible to write:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | Our results coincide with this picture from the Bishop Book, showing that much regularization leads to a worse fit (higher bias) but less variance.l |